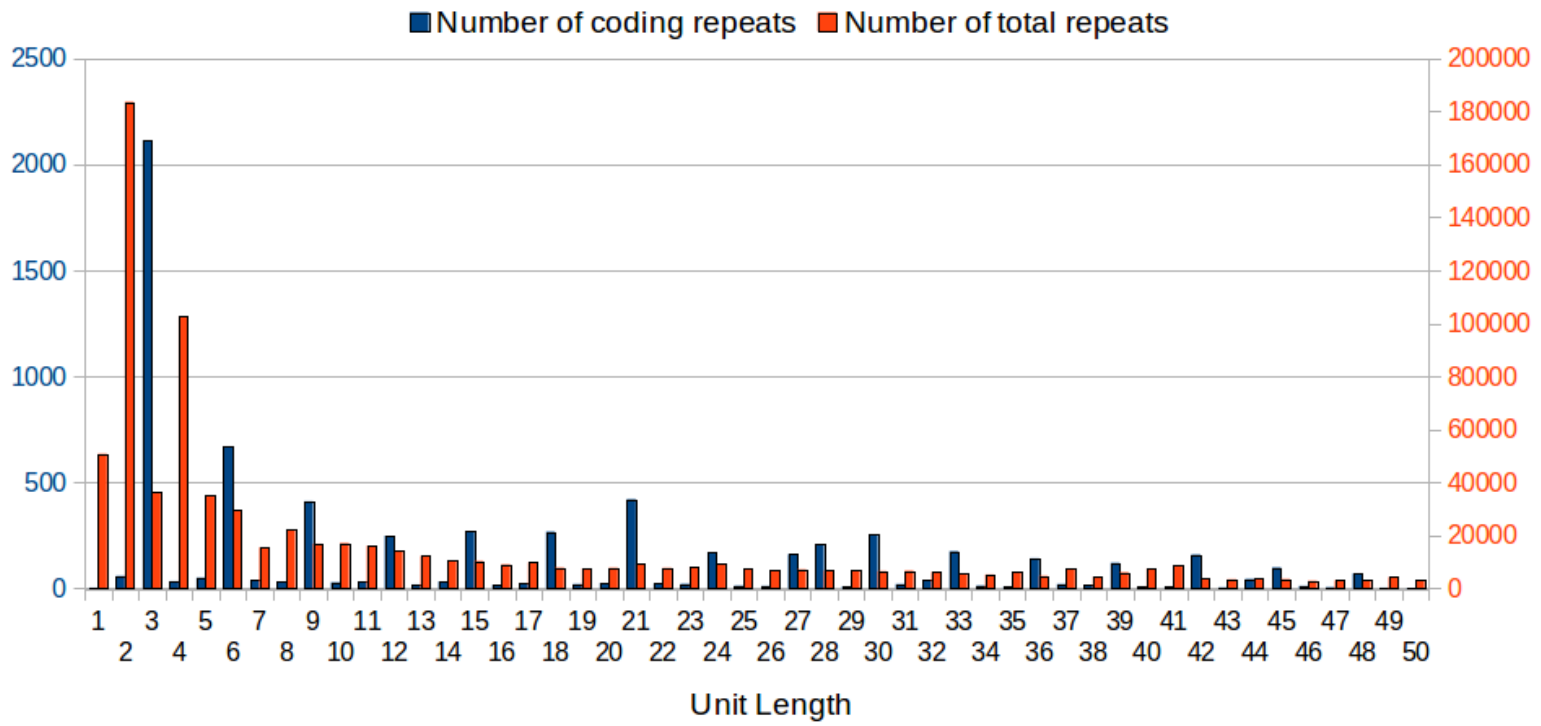
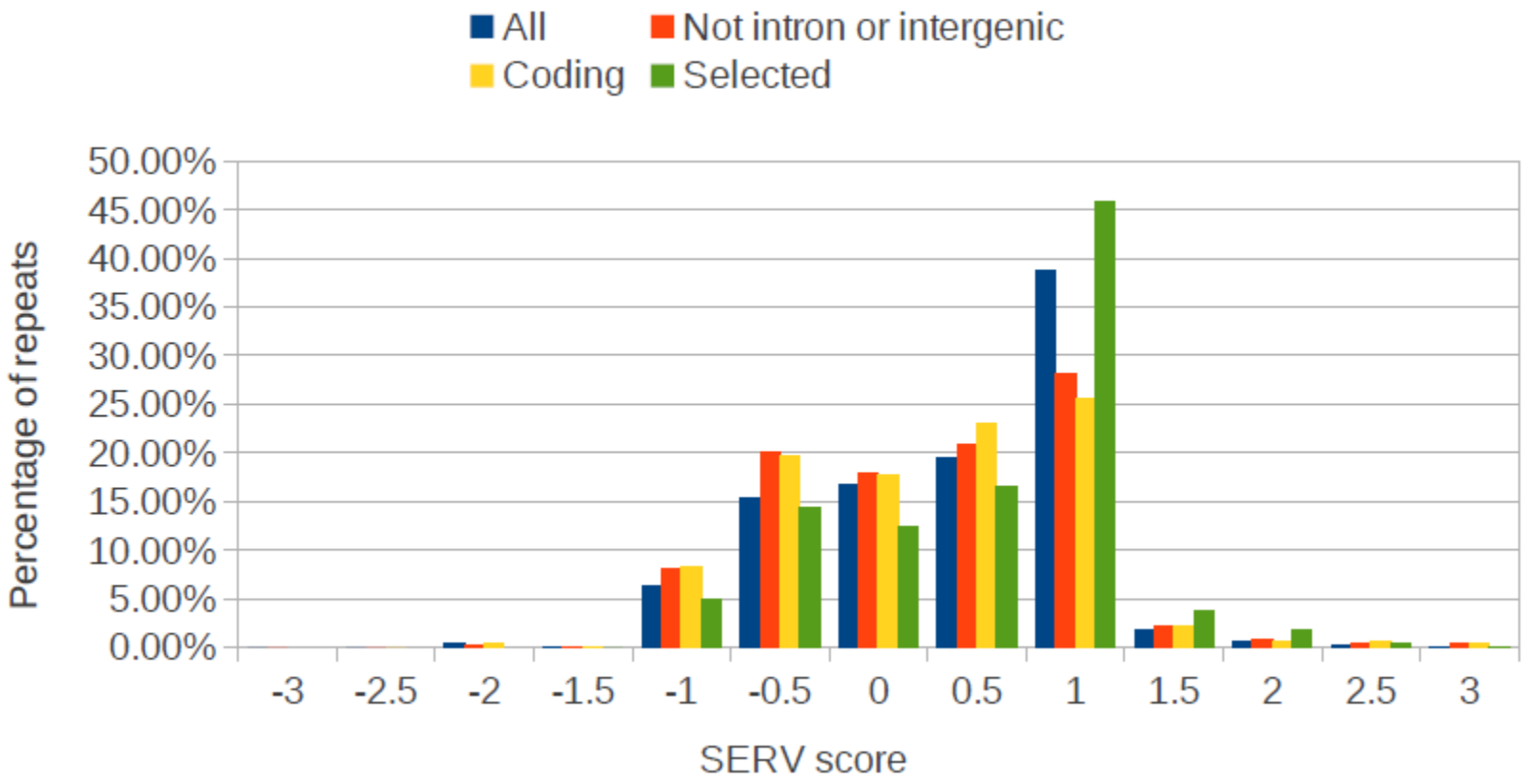
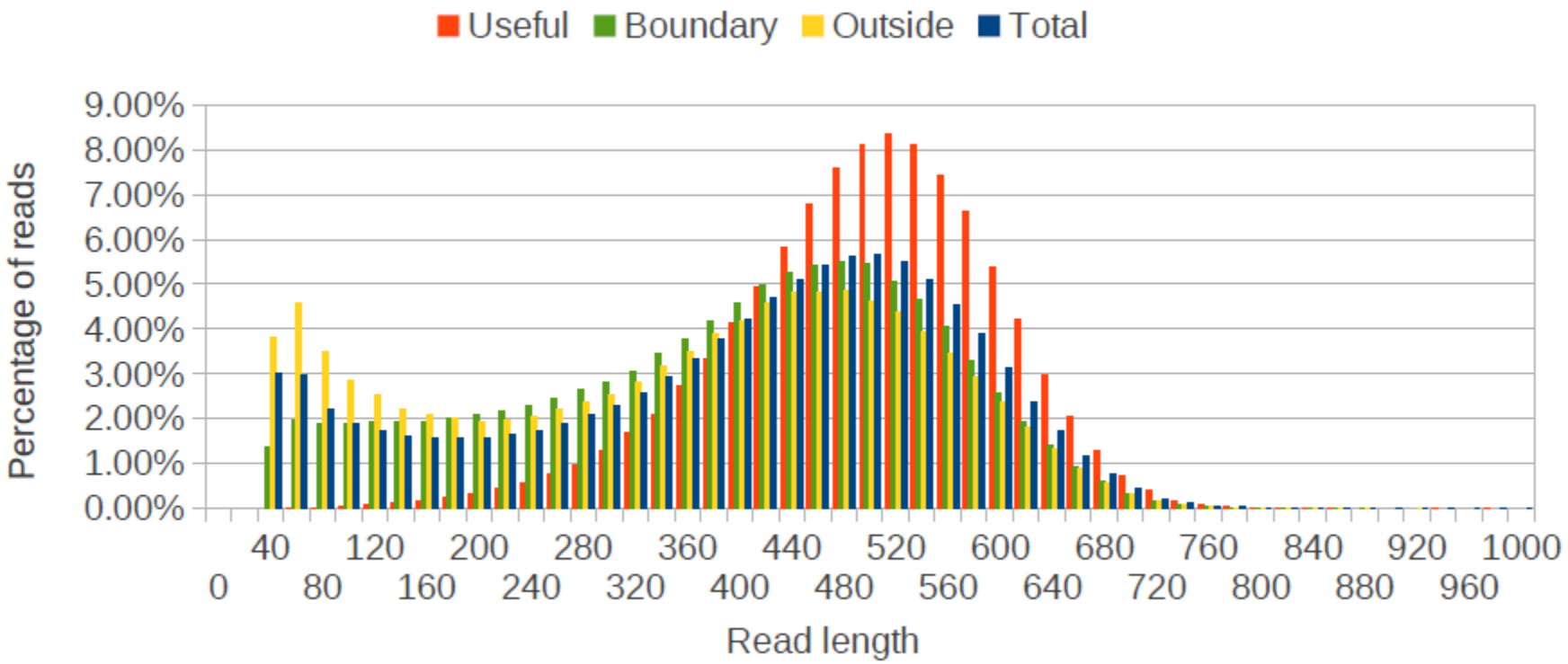


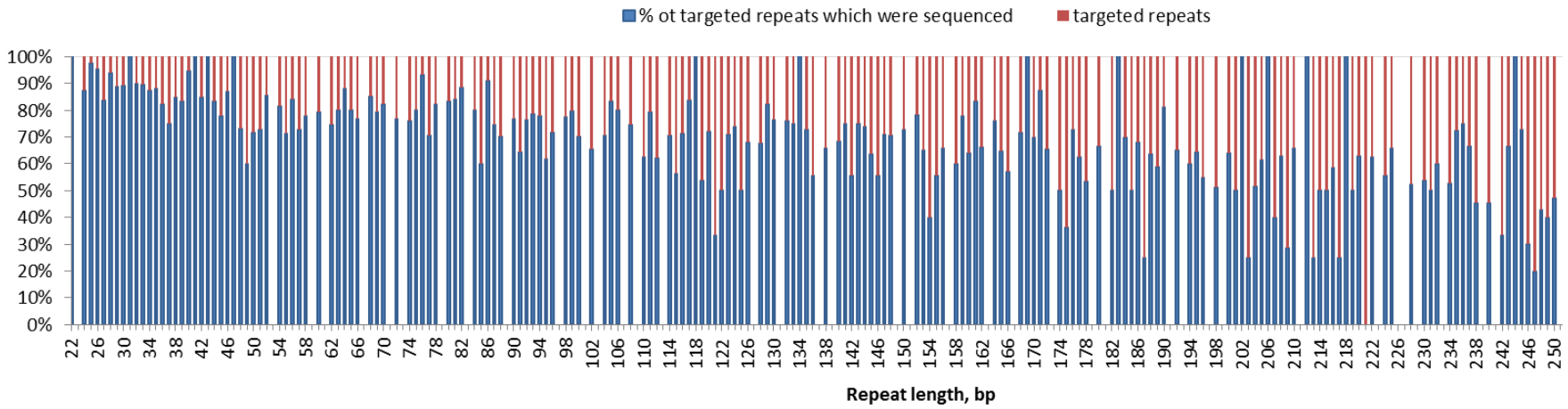
Supplemental Figure S1. Distribution of repeats for different unit lengths.



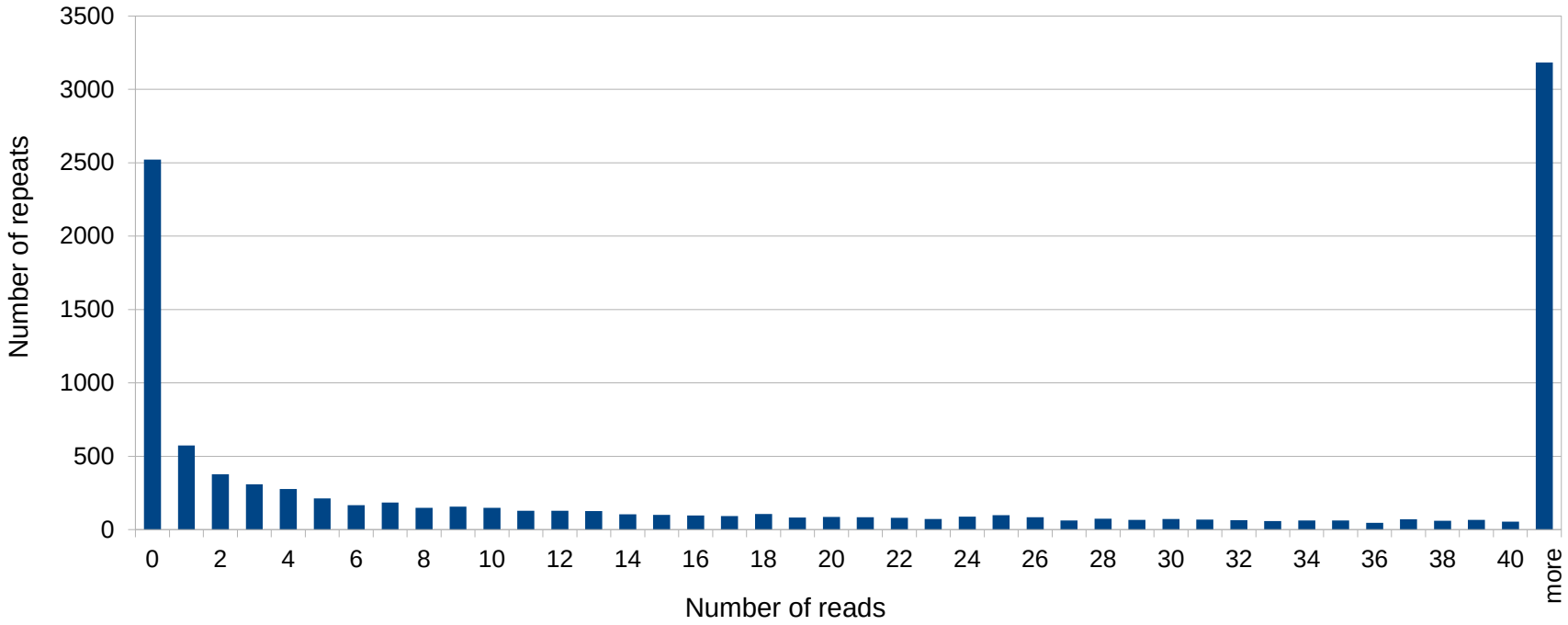
Supplemental Figure S2. Histogram of SERV scores for the catalog of repeats. Repeats in coding regions are predicted to be less variable.



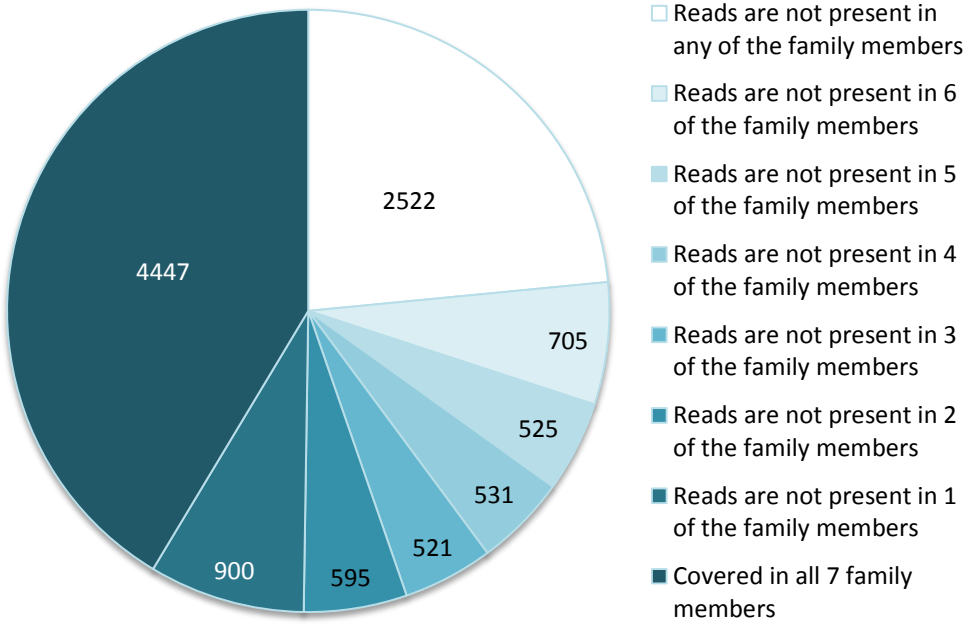
Supplemental Figure S3. Histogram of read lengths obtained after targeted sequencing of 7 individuals of a family. Useful: 338,046 reads that span the entire repeat and contain at least 45 bp of both flanking sequences; Boundary: 526,302 reads located within the flanking sequences but not spanning the repeat; Outside: 478,982 reads that map to other non-captured regions; Total: 1,400,540 reads. Useful repeats are found predominantly for reads with lengths >120 bp.



Supplemental Figure S4. Percentage of targeted repeats (indicated in blue) that were sequenced in the seven individuals relative to their repeat lengths.



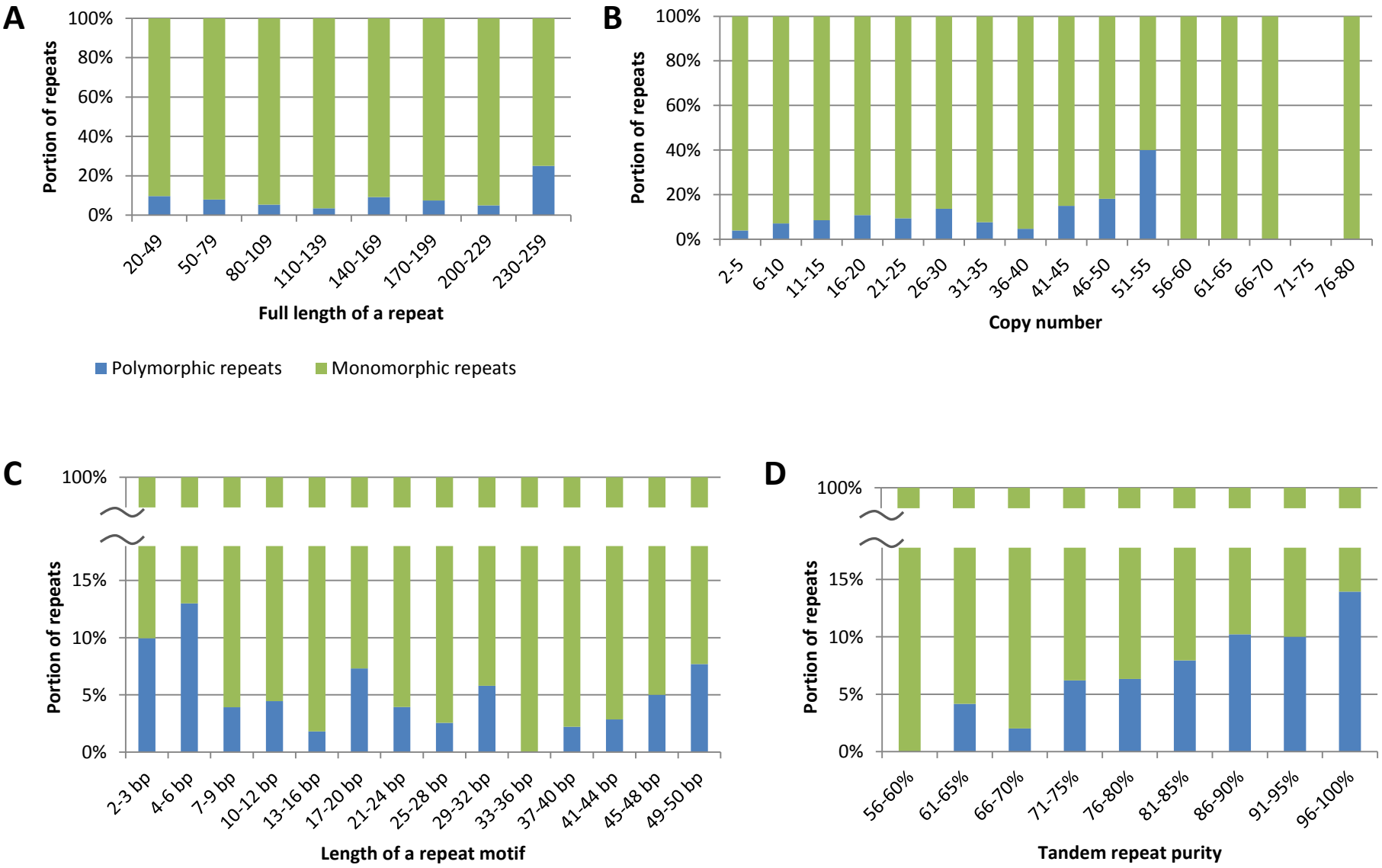
Supplemental Figure S5. Distribution of number of repeats for different coverage levels, calculated as the total read counts over the seven individuals.



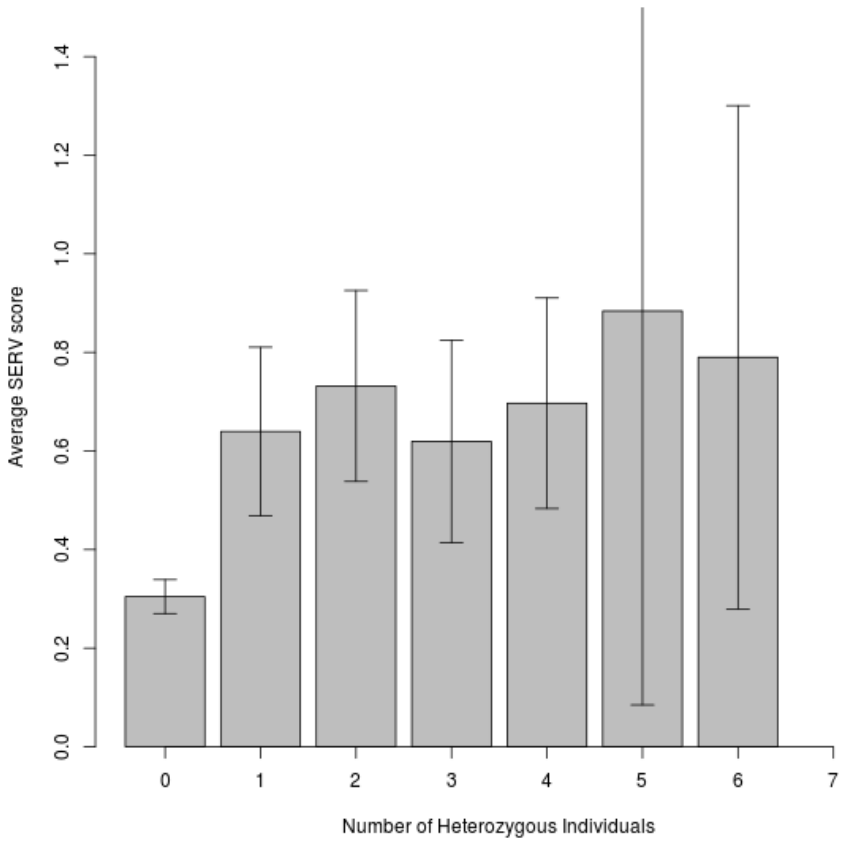
Supplemental Figure S6. Numbers of covered repeats in 0 to 7 individuals.

Locus					Sequencing-based data				Fragment analysis-based genotype call
Chromo-some	Start	End	Unit length	Unit sequence	Allele length	Copy number	Number of reads	Genotype call	
chr11	83231586	83231781	4 bp	ATAG	196 bp 236 bp 240 bp	C1 = 49 C2 = 59 C3 = 60	R1 = 5 R2 = 2 R3 = 1*	49|59	49|60

Supplemental Figure S7. Example of a missed allele due to the calling of its stutter. The indicated repeat at chromosome 11 in grandfather (GF) obtained the GS-FLX+ genotype calls 49 and 59 as the ATAG unit numbers. These figures were based on the 5 reads with 49 units (GF1-5) and 2 reads of the 59 units (GF6-7; shown below). *****The single read of the correct 60 unit allele (GF8; also indicated below) was therefore not called since only the alleles with the highest read numbers are used. Fragment analysis demonstrated however, that the true alleles are 49 and 60 showing that the stutter (59) of the 60 allele was incorrectly called here. Note that this repeat consist of two tandem repetitive ATAG parts (horizontal arrows), which are both variable. The stutter occurs at the first (longest) part.



Supplementary Figure S8. Distribution of polymorphic and monomorphic repeats for **A.** repeat length; **B.** copy number; **C.** unit (motif) length; **D.** repeat purity.



Supplementary Figure S9. Average SERV scores for repeats with different numbers of heterozygous individuals. Conserved repeats have significantly smaller SERV scores.